

yang_seonhyeHW24

```
library(faraway)
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
bs <- read.delim("/cloud/project/bacterial-soap.txt")
class(bs$Method)

## [1] "factor"
##Question 1.a
#subsetting data into each method and calculating the mean and sd for bacterial count
Water=bs[bs$Method %in% c("Water"), ]
Soap=bs[bs$Method %in% c("Soap"), ]
AlcoholSpray=bs[bs$Method %in% c("AlcoholSpray"), ]
AntibacterialSoap=bs[bs$Method %in% c("AntibacterialSoap"), ]

#mean
mean(Water$BacterialCounts)

## [1] 117
mean(Soap$BacterialCounts)

## [1] 106
mean(AlcoholSpray$BacterialCounts)

## [1] 37.5
mean(AntibacterialSoap$BacterialCounts)

## [1] 92.5
#standard deviation
sd(Water$BacterialCounts)

## [1] 31.13106
sd(Soap$BacterialCounts)

## [1] 46.95895
sd(AlcoholSpray$BacterialCounts)

## [1] 26.55991
sd(AntibacterialSoap$BacterialCounts)
```

```
## [1] 41.96257
```

Question 1.b

```
fit <- lm(BacterialCounts~Method, data = bs)
null <- lm(BacterialCounts~1, data = bs)

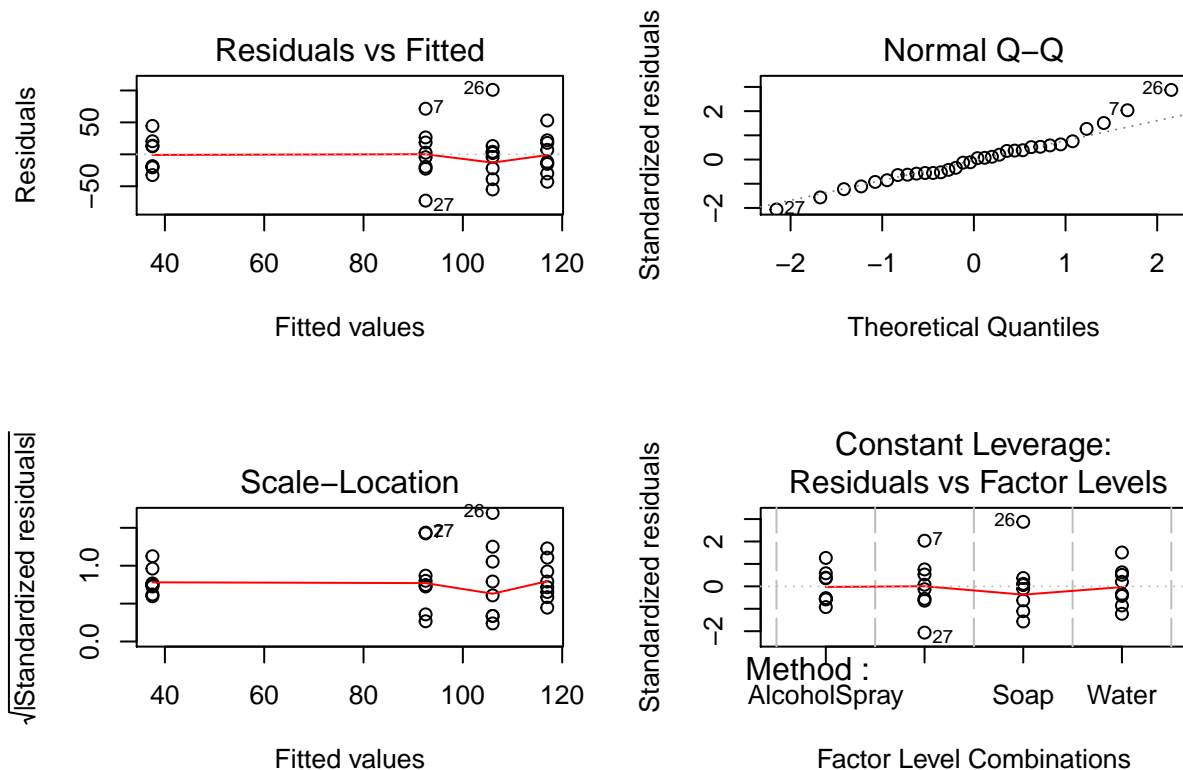
anova(null, fit)
```

```
## Analysis of Variance Table
##
## Model 1: BacterialCounts ~ 1
## Model 2: BacterialCounts ~ Method
##   Res.Df  RSS Df Sum of Sq    F   Pr(>F)
## 1      31 69366
## 2      28 39484   3     29882 7.0636 0.001111 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is 0.001111 and since the p-value is less than alpha, we reject the null hypothesis. Therefore, we can conclude that the mean of BacterialCounts for all Methods is not equal.

Question 1.c

```
par(mfrow=c(2, 2))
plot(fit)
```



There seems to be a linear relationship by looking at the Residuals vs Fitted (constant variance is satisfied),

there seems to be a normal distribution, the residuals are spread equally along the ranges of predictors which implies that the assumption of equal variance (homoscedasticity) is satisfied and by looking at the Constant Leverage plot, there are no signs of outliers.

Question 1.d

```
rstudent(fit)

##           1           2           3           4           5           6
## -1.23560535 -0.61937483 -0.63365851  0.37839865  0.50557528 -1.60963481
##           7           8           9          10          11          12
##  2.16544263 -0.92276918 -0.42070469  0.11184791 -0.12583664 -0.51975764
##          13          14          15          16          17          18
##  0.19582743 -1.11508671  0.51975764 -0.54815720 -0.33616733  0.36431366
##          19          20          21          22          23          24
## -0.54815720  0.57660544  0.61937483  0.05591424  0.74846669  0.35023666
##          25          26          27          28          29          30
##  1.54581569  3.36338102 -2.20111900  1.28128187 -0.84980696 -0.11184791
##          31          32
##  0.06989508 -0.57660544

critval <- qt(0.05/(2*nobs(fit)), df=df.residual(fit)-1, lower=FALSE)
which(abs(rstudent(fit)) > critval)

## named integer(0)
```

According to Bonferroni test, there are no outliers.

Question 1.e

```
TukeyHSD(aov(fit, data=bs))

##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = fit, data = bs)
##
## $Method
##
##           diff           lwr           upr           p adj
## AntibacterialSoap-AlcoholSpray 55.0    3.735849 106.26415 0.0319648
## Soap-AlcoholSpray              68.5   17.235849 119.76415 0.0055672
## Water-AlcoholSpray             79.5   28.235849 130.76415 0.0012122
## Soap-AntibacterialSoap         13.5  -37.764151  64.76415 0.8886944
## Water-AntibacterialSoap        24.5  -26.764151  75.76415 0.5675942
## Water-Soap                    11.0  -40.264151  62.26415 0.9355196
```

Looking at the p-values from our results, we can see that AntibacterialSoap-AlcoholSpray, Soap-AlcoholSpray and Water-AlcoholSpray are significant because their p-values are less than 0.05 and their ranges don't include 0.